4. Topological phases-b

Dirac fermions

Two-band system

We first calculate the Berry phase for one- and two-dimensional Hamiltonians. For the sake of simplicity, we shall focus on the discussion of two-band systems in two dimensions, which are the relevant cases to compare with the Dirac theory.

The simplest insulator possesses two bands, one above and one below the band gap. Such an insulator can generically be described as a two-level system, which corresponds to a two-dimensional Hilbert space at each point of the Brillouin torus, on which acts a Bloch Hamiltonian continuously defined on the Brillouin torus. That is, in two dimensions, the model for a ***continuum*** Dirac fermion is a variant of the general two-band Hamiltonian, which can be written as a 2x2 Hermitian matrix, parameterized by the real functions h_μ:

$$H(k) = h_\mu(k)\sigma_\mu = h_0(k)\mathbb{I} + \vec{h}(k)\cdot\vec{\sigma} = \begin{pmatrix} h_0 + h_3 & h_1 - ih_2 \\ h_1 + ih_2 & h_0 - h_3 \end{pmatrix}$$

where $\vec{h}(k) = (h_1, h_2, h_3)$ is an arbitrary three-vector and $\vec{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ are the Pauli matrix. Notice that here the momentum **k** is a continuum variable lying in the infinite Euclidean plane.

In the following, we always assumed that the coefficients h_μ are well defined on Brillouin torus, i.e. are periodic. The spectral theorem ensures that $H(k)$ has two orthogonal normalized eigenvectors $|\pm\rangle$ with eigenvalues $E_\pm$. The eigenvectors are such that $\mathrm{H}|\pm\rangle = E_\pm|\pm\rangle$, $\langle+|+\rangle = 1$, $\langle-|-\rangle = 1$, $\langle+|-\rangle = 0$, $\langle-|+\rangle = 0$. Using $\mathrm{Tr}(H) = 2h_0$ and $\det(H) = h_0^2 - h^2$ with $h(k) = \left\|\vec{h}(k)\right\| = \sqrt{h_1^2 + h_2^2 + h_3^2}$. The two eigenvalues are $E_\pm = h_0(k) \pm h(k)$.

The corresponding normalized eigenvectors are, up to a phase:

$$|\pm\rangle = \left(1 + \frac{h_3 + E_\pm^2}{h_1^2 + h_2^2}\right)^{-1/2} \begin{pmatrix} \frac{E_\pm}{h_1 + ih_2} \\ 1 \end{pmatrix}$$

The energy shift of both energies has no effect on topological properties, provided the system remains insulating. To simplify the discussion, let us take $h_0 = 0$. Therefore, the system is insulating provided $h(k)$ never vanishes on the whole Brillouin torus, which we enforce in the following. As we focus only on the topological behavior of the filled band, which is now well-defined, we only consider the filled eigenvector $|-\rangle$ in the following.

A. The One Dimensional Case

For one dimensional systems, we shall discuss two topological invariants, one is the Berry phase for off-diagonal Hamiltonians ($h_3 = 0$), and the other is the Majorana number M for a Hamiltonian with $h_3 \neq 0$. Let us first discuss the Berry phase.

We consider $h_3 = 0$ in the Hamiltonian and $|\pm\rangle$. Thereby, the eigenstates are

$$|\pm\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} \pm 1 \\ \hat{h}^*(k) \end{pmatrix}, \quad \hat{h}(k) = \frac{h(k)}{|h(k)|}$$

The Berry phase is, for the valence band $E = E_-$, is given by

$$\gamma(E_-) = -\int_{B.Z.} dk\, \mathcal{A}(k) = -2i \int_{B.Z.} \langle -|\partial_k|-\rangle dk = -i \int_{B.Z.} dk \left[\hat{h}^*(k)\right]^{-1} \partial_k \hat{h}^*(k)$$

where the extra factor 2 is due to the two spins up and down. The minus sign is due to the general relation $\gamma(E_-) = -\gamma(E_+)$, which shows that the valence band has opposite sign in comparison with the conduction band. Using $dh^* = dk\partial h/\partial k$, $\gamma(E_-)$ becomes

$$\gamma(E_-) = \frac{1}{i}\oint_{C_h} \frac{d\hat{h}^*}{\hat{h}^*} = \frac{1}{i}\oint_{C_h} \frac{dh^*}{h^*}$$

where C_h is a closed path in the h-space. This is because the integral in k is over the first Brillouin zone with $h(k_x + 2\pi, k_y + 2\pi) = h(k_x, k_y)$. The topological invariant $\gamma(E_-)$ is conventionally called Zak phase instead of Berry phase in 1D.

The expression of $\gamma(E_-)$ has a geometrical interpretation. Indeed, let us consider the winding number ω of a closed curve C in the complex plane around a point a. In complex analysis, this is given by

$$\omega = \frac{1}{2\pi i}\oint_C \frac{dh}{h - a_0}$$

where $h = h_x + ih_y = r\exp(i\theta)$ is a complex number. By comparing the expressions between $\gamma(E_-)$ and ω, we conclude that: **The winding number is the Berry phase divided by 2π**.

Next let us consider $h_3 \neq 0$. In the basis $\mathcal{C}_k^\dagger = \begin{pmatrix} c_k & c_{-k}^\dagger \end{pmatrix}$, the Hamiltonian reads $H = \mathcal{C}_k^\dagger h(k) \mathcal{C}_k$. We define a new basis given by

$$a_k = c_k + c_{-k}^\dagger, \quad ib_k = c_k - c_{-k}^\dagger$$

After some calculation, one may show that the Hamiltonian in this basis is given by $H = 1/2 A^\dagger B(k) A$, where $A^\dagger = \begin{pmatrix} a_k & b_k \end{pmatrix}$, $B(k) = \mathbf{b}\cdot\sigma$, and $\mathbf{b} = (-h_2, -h_3, h_1)$. The Majorana number M is defined as the sign of the product of the Pfaffian of $B(k)$ at $k = \pi$ and $k = 0$, i.e. $M = \text{sign}\{\text{Pf}[B(k=\pi)]\}\text{sign}\{\text{Pf}[B(k=0)]\}$. The Pfaffian of $B(k)$ is $\{\text{Pf}[B(k)]\} = \pm\sqrt{\det B(k)} = \text{sign}(E_\pm)\left|E_\pm(k)\right|$. Therefore,

$$M = \text{sign}[h_3(k=\pi)]\text{sign}[h_3(k=0)]$$

In this expression, we have used that the sign of h_3 is the sign of the energy band. This is because the z-axis is the privileged axis within our convention for the Pauli matrix. The real operators (a_k, b_k) are called Majorana operators. They may always be obtained independent on the system we are considering, because $a_k = c_k + c_{-k}^\dagger$, $ib_k = c_k - c_{-k}^\dagger$ is a generic transform. Hence, the Majorana number is essential to distinguish between a trivial or a Majorana phases. For $M = +1$ the phase is trivial, but $M = -1$ there exist Majorana phase, which must be achieved by tuning the parameters of the Hamiltonian. We shall apply this concept for the Kitaev Chain later.

B. The Two Dimensional Case

In order to distinguish from the 1D case, we now use $\mathrm{d_i}$ to replace h_i for the 2D case. Specifically, Dirac fermions in two dimensions are described by the Hamiltonian

$$H(\boldsymbol{k}) = \sum_i d_i(\boldsymbol{k}) \cdot \sigma_i \quad with \quad d_1(\boldsymbol{k}) = k_x, \quad d_2(\boldsymbol{k}) = k_y, \quad d_3(\boldsymbol{k}) = m \tag{0.1}$$

which is called Dirac Hamiltonian. The energies and eigenstates are given by

$$\epsilon(\mathbf{k})_\pm = d_\pm(\mathbf{k}) = \pm\sqrt{k^2 + m^2}$$

$$\psi_\pm(\mathbf{k}) = \frac{1}{\sqrt{2d(\mathbf{k})[d(\mathbf{k}) \pm d_3(\mathbf{k})]}} \begin{pmatrix} d_3(\mathbf{k}) \pm d(\mathbf{k}) \\ d_1(\mathbf{k}) - id_2(\mathbf{k}) \end{pmatrix}$$

It is straight forward to show that the Berry connection of the lower band can be written as

$$\mathcal{A}_\mu(\boldsymbol{k}) = i\left\langle \psi_-(\boldsymbol{k}) \middle| \partial_{k_\mu} \middle| \psi_-(\boldsymbol{k}) \right\rangle = -\frac{1}{2d(\boldsymbol{k})[d(\boldsymbol{k}) + d_3(\boldsymbol{k})]} \left[d_2(\boldsymbol{k}) \partial_{k_\mu} d_1(\boldsymbol{k}) - d_1(\boldsymbol{k}) \partial_{k_\mu} d_2(\boldsymbol{k}) \right] \tag{0.2}$$

And the corresponding Berry curvature is given by

$$\mathcal{F}_{\mu\nu}(\boldsymbol{k}) = \frac{1}{2d^3} \epsilon_{\alpha\beta\gamma} \hat{d}_\alpha(\boldsymbol{k}) \partial_{k_\mu} \hat{d}_\beta(\boldsymbol{k}) \partial_{k_\nu} \hat{d}_\gamma(\boldsymbol{k}) \quad with \quad \hat{\boldsymbol{d}}(\boldsymbol{k}) = \frac{\boldsymbol{d}(\boldsymbol{k})}{|d(\boldsymbol{k})|} \tag{0.3}$$

Hence, using our concrete **d**-vector we find

$$A_x = \frac{-k_y}{2\sqrt{k^2+m^2}\left(\sqrt{k^2+m^2}+m\right)} \quad and \quad A_y = \frac{k_x}{2\sqrt{k^2+m^2}\left(\sqrt{k^2+m^2}+m\right)} \tag{0.4}$$

and therefore the Berry curvature

$$\mathcal{F}_{xy} = \frac{m}{2(k^2+m^2)^{3/2}} \tag{0.5}$$

To calculate the specific form of the Berry phase, let's assume we are at finite chemical potential μ larger than the mass gap m, which endows us with a Fermi surface. Note that although we did not implicitly put the chemical potential into the Hamiltonian, the Berry connection remains unchanged because the chemical potential is a diagonal identity matrix term and the eigenstates are not changed by diagonal terms. If the eigenstates are unchanged, the Berry connection is

unchanged as well. We then would like to integrate the Berry curvature over the Fermi surface, which is an azimuthal integration over the angle at a momentum $k_F = \sqrt{\mu^2 - m^2}$. For the integral of the Berry connection over the Fermi surface, we obtain the Berry phase

$$\int_{Fermi\ Surface} d\mathbf{k}\cdot\mathbf{A} = \int_0^{2\pi} k_{\mathrm{F}} d\theta \mathcal{A}_\theta = \int_0^{2\pi} k_{\mathrm{F}} d\theta \left(\mathcal{A}_y \frac{k_x}{k_F} - \mathcal{A}_x \frac{k_y}{k_F}\right) = \int_0^{2\pi} \frac{k_F^2}{2\sqrt{k_F^2+m^2}\left(\sqrt{k_F^2+m^2}+m\right)} d\theta$$

$$= \pi \frac{k_F^2}{\sqrt{k_F^2+m^2}\left(\sqrt{k_F^2+m^2}+m\right)} \approx \pi - \pi\frac{m}{k_{\mathrm{F}}}$$

Notice that the Berry phase at the TR-invariant (gapless) $m = 0$ point is π; this be can easily understood because π is the only value (besides 0) that is TR invariant (modulo 2π). When time reversal is broken, the Berry phase around the Fermi surface picks up a contribution proportional to the ratio between the gap and the Fermi momentum—so if the Fermi level is really high, corresponding to large doping, then the Berry phase can be again very close to (but not equal) π , even though the time reversal is broken.

Similar to the one dimensional problem, we define a topological invariant C, the Chern number

$$C = \frac{1}{4\pi}\int_{BZ} \hat{d}\cdot\left(\frac{\partial\hat{d}}{\partial k_x}\times\frac{\partial\hat{d}}{\partial k_y}\right) dk_x dk_y = \frac{1}{4\pi}\int_0^\infty\int_0^{2\pi} dk d\phi \frac{km}{(k^2+m^2)^{\frac{3}{2}}} = \frac{sign(m)}{2}$$

Let us put the chemical potential in the gap ($|\mu| <$ m) and integrate the Berry curvature over the occupied states (this is the Berry curvature of the occupied band), we then have the Hall conductance

$$\sigma_{xy} = \frac{e^2}{h}\frac{1}{2\pi}\int dk\mathcal{F}_{xy} = \frac{e^2}{h}\int_0^\infty dk \frac{mk}{2(k^2+m^2)^{3/2}} = \frac{e^2}{2h} sign(m) \qquad (0.6)$$

We have arrived at the notable result that massive 2D Dirac fermions have a Hall conductance equal to one-half times the sign of their mass. This statement is puzzling because the Hall conductance, equal to the Chern (winding) number of the filled band, **has to be an integer**.

With that, we can draw several important insights from this results:

1. **$\sigma_{xy} \neq 0$, we must have broken time-reversal invariance. How did this happen?**
2. **$\sigma_{xy} \neq \frac{e^2}{h}\nu$ with $\nu \in Z$. How can this be?**

Answer to 1: Let us start with the first question. We have to make the distinction between two cases. (i) If the σ-matrices encode a real spin-1/2 degree of freedom (d_i correspond to real spin), the time reversal operator is given by

$$\mathcal{T} = -i\sigma_y K$$

where K denotes complex conjugation. This is actually the definition of time reversal operator defined in a two-band model. all three matrices σ_y are odd under time reversal. Therefore

$$\mathcal{T}H(\boldsymbol{k})\mathcal{T}^{-1} = \sum_i -d_i(\boldsymbol{k})\,\sigma_i = -k_x\sigma_x - k_y\sigma_y - m\sigma_z$$

If we want to above Hamiltonian to be time reversal invariant we need to have $[H(\boldsymbol{r}), \mathcal{T}\,] = 0$, or in the momentum representation, this has to be

$$\mathcal{T}H(\boldsymbol{k})\mathcal{T}^{-1} = H(-\boldsymbol{k}) = -k_x\sigma_x - k_y\sigma_y + m\sigma_z$$

From this we conclude that the Dirac fermions are only time reversal invariant for $d_3(\boldsymbol{k}) = m = 0$. However, in this case, there is **no gap** in the spectrum at $\boldsymbol{k} = 0$ and the calculation of σ_{xy} does not make sense. Otherwise, the Dirac Hamiltonian is not time reversal invariant.

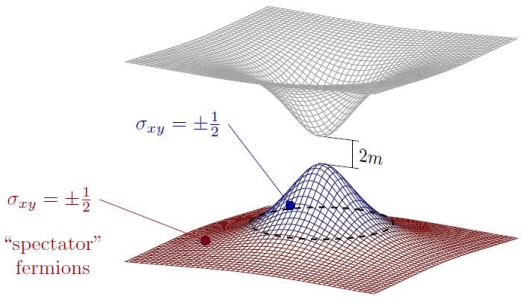
(ii) For the case that the Pauli matrices describe some iso-spin (d_i corresponds to iso-spin), where $\mathcal{T} = K$, we need to have $\mathcal{T}^2 = 1$ and $H(\boldsymbol{k}) = H(-\boldsymbol{k})$. This implies that

$$d_1(\boldsymbol{k}) = d_1(-\boldsymbol{k}),\;\; d_2(\boldsymbol{k}) = -d_2(-\boldsymbol{k}),\;\; d_3(\boldsymbol{k}) = d_3(-\boldsymbol{k})$$

It then turns out that to maintain time reversal, we need d_1, d_3 to be even in $\boldsymbol{k}$ but d_2 to be odd. Of course, in this case the Chern number has to vanish.

Answer to 2: Let us now address the noninteger-quantized nature of σ_{xy}. The quantization of σ_{xy} arises from the quantized value of the Chern number. We have seen in our derivation, however, that it was crucial that the domain over which we integrated the Berry curvature was closed and orientable. Here we are in a continuum model where the integral over all momenta extends over the whole $\mathbb{R}^2$ (2D real coordinate space). We have therefore no reason to expect σ_{xy} to be quantized.

Imagine that the Dirac Hamiltonian arises from some low-energy expansion ($\mathbf{k}\cdot\mathbf{p}$) around a special point in the Brillouin zone of a lattice model. For the full lattice, the $k \to \infty$ integral would be regularized due to the Brillouin zone boundary. The whole system has a quantized Hall conductivity. However, the region close to the "Dirac point" contributes $\pm 1/2$ to the Chern number. Moreover, imagine a gap closing and re-opening transition described by the Dirac Hamiltonian where m changes its value. In such a situation the change in Chern number $\Delta\mathcal{C} = \pm 1$.



The reason we get one-half is because the Dirac fermion is in the continuum and has not been regularized properly. The bands of the continuum Dirac fermion do not have a bandwidth. The Hall conductance equals the Chern number and is an integer only if the base manifold (the BZ) is compact. In the continuum, the momentum runs over a noncompact manifold (the infinite Euclidean plane), and this does not apply.

How does having a lattice fix things? If we are on a lattice, the bands, which in the case of the Dirac fermion have infinite bandwidth (because the dispersion is $\pm\sqrt{k^2+m^2}$), must bend down due to the fact that the problem has a finite bandwidth. At the points where the bands bend down, we will get another half a quantum of Hall conductance. These high energy fermions, which contribute to conductances, are called spectator fermions.

In an insulator, the vicinity of the *k*-space point where the gap is the smallest can be modeled by a Dirac Hamiltonian. However, due to the presence of the spectator fermions just mentioned, we cannot calculate the Hall conductance of the filled lattice band just by looking at the points where the gap is the smallest and then performing the Dirac computation. The reason is that we might have some higher energy modes that can add or subtract half-values.

However, the change in Hall conductance upon closing or reopening the gap can be determined by focusing only on the vicinity of the point where the transition happens. Generically, a gap-closing-and-opening transition has to happen by varying a parameter, the mass of the Dirac fermion *m*. If we go from *m* negative to *m* positive, then we get the Hall conductance to change by 1. If we knew the value of the Hall conductance before the gap-closing transition, we would be able to find the value of the Hall conductance after closing and reopening the gap.

The correct procedure to calculate the Chern number is to perform the calculations in the lattice theory, after that to apply the desired limits. It would be emphasized that this has been done in A. Coste &M. Luscher, Parity Anomaly and Fermion- Boson Transmutation in 3-Dimensional Lattice QED, Nucl. Phys. B323, 631 (1989) for the Dirac theory. Indeed, the authors have shown that, by using the lattice parameter a as a natural regulator, one obtains a quantized Hall conductivity, using the Kubo formula with the vacuum polarization tensor that they have calculated. Therefore, the quantization of the Hall conductivity is related to the lattice regularization of the parity anomaly. In the massless case $m = 0$, it yields a spontaneous quantum valley Hall effect (E. C.Marino, L. O. Nascimento, V. S. Alves, & C.Morais Smith, Interaction Induced Quantum Valley Hall Effect in Graphene, Phys. Rev. X 5, 011040 (2015)). In this case, this valley conductivity is quantized exact as the Hall conductivity for graphene $\sigma_{V,xy} = 4(n+1/2)e^2/h$ and Landau-like energy levels are dynamically generated due to the electromagnetic interaction. The result shown by A. Coste et. al. solves this long misunderstand between anomalous quantum Hall effect and parity anomaly. Indeed, using their result, a well-quantized Hall conductivity emerges from the Kubo formula.

***The Berry Phase for a Generic Dirac Hamiltonian in Two Dimensions**

In the previous derivation of the Berry phase, we have assumed rotational invariance. However, around a degeneracy point or band, crossing in two-dimensions, we have the following general form for the Hamiltonian:

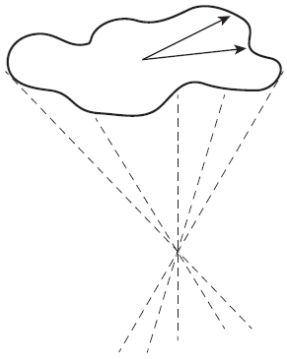
$$h(\mathbf{k}) = A_{ij}k_i\sigma_j$$

with A_{ij} a 2×2 matrix. We have neglected a possible diagonal term proportional to the identity matrix because it does not influence the eigenstates and, hence, does not enter the Berry phase.

The eigenvalue of the Hamiltonian $d = \sqrt{(A_{1i}k_i)^2 + (A_{2i}k_i)^2}$ and the Berry connection is

$$A_{1i}(\mathbf{k}) = -\frac{1}{2d^2}\left(A_{2j}k_jA_{1i} - A_{1j}k_jA_{2i}\right)$$

We want to compute the Berry phase over the Fermi surface (see figure), but, in general, the Fermi surface can be anisotropic. However, we have $A_i dk_i = A_\phi k(\phi)d\phi$ because an infinitesimal variation $d\mathbf{k} = k(\phi)d\phi\hat{1}_\phi$ on the Fermi surface. Here, $\hat{1}_\phi$ is the unit vector tangential to a surface of equal energy in *k*-space. We have that



Fermi surface for generic, noncircular Dirac Hamiltonian.

$$A_\phi = A_y\frac{k_x}{k} - A_x\frac{k_y}{k}$$

so the Berry phase is

$$\gamma = \int_0^{2\pi} d\phi\left(A_yk_x - A_xk_y\right) = \int_0^{2\pi} d\phi\,\frac{1}{2d^2}k^2\det A$$

which, upon substituting for d and using the integral

$$\int_0^{2\pi} d\phi\,\frac{1}{a\cos(\phi)^2+b\sin(\phi)^2+c\cos(\phi)\sin(\phi)} = \frac{4\pi}{\sqrt{4ab-c^2}}$$

valid as long as $c^2 < 4ab$, which is the condition of not having nodal lines, brings us to a nice expression for the Berry phase:

$$\int_{Fermi\ Surface} d\mathbf{k}\cdot\mathbf{A} = \pi\cdot\mathrm{sign}(\det A)$$

Edge states in the lattice Chern insulator

Topological materials have an important property: their surface states are stable against perturbations. They can be destroyed only if the energy gap of the bulk bands closes so that the topology of the electronic states is trivialized. In general, the interface between two materials with different topologies would have robust interface states.

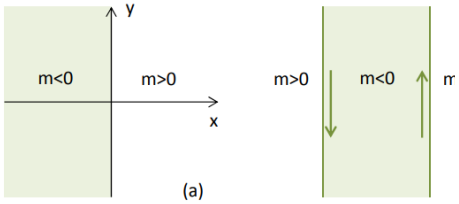
Take the 2D lattice model as an example. Divide the space to two parts where

$$M(x)\begin{cases} > 0 \text{ for } x > 0 \\ < 0 \text{ for } x < 0 \end{cases}$$

so that there is a 1D boundary along the *y*-axis (see Fig. 40(a)). For simplicity, consider only the small **k** limit,

$$H = \varepsilon_0 + \begin{pmatrix} M & \lambda(k_x - ik_y) \\ \lambda(k_x + ik_y) & -M \end{pmatrix} + O(k^2)$$

The exact profile of $M(x)$ does not matter, as long as it is monotonic and smooth (compared to the electron wavelength λ). To solve for the surface states, one needs to re-quantize the Hamiltonian using the substitution $k \to \frac{\partial}{i\partial \mathbf{r}}$, such that



$$H(\mathbf{p}) = \varepsilon_0 + \begin{pmatrix} M(x) & \lambda\left(\frac{1}{i}\frac{\partial}{\partial x} - \frac{\partial}{\partial y}\right) \\ \lambda\left(\frac{1}{i}\frac{\partial}{\partial x} + \frac{\partial}{\partial y}\right) & -M(x) \end{pmatrix}$$

The *x*-directions extend to infinity on both ends, and the periodic boundary condition is imposed along the *y*-direction. We now solve the differential equation,

$$H(\mathbf{p})\psi(x,y) = \varepsilon\psi(x,y)$$

Use the method of separation of variables and write

$$\psi(x,y) = \phi_1(x)\phi_2(y)$$

Since the *y*-direction is invariant under translation, a trivial solution is $\phi_2(y) = e^{ik_y y}$, a plane wave. Therefore, the equation for $\phi_1(x)$ is,

$$\begin{pmatrix} M(x) & \frac{\lambda}{i}\left(\frac{\partial}{\partial x} + k_y\right) \\ \frac{\lambda}{i}\left(\frac{\partial}{\partial x} - k_y\right) & -M(x) \end{pmatrix}\phi_1(x) = \varepsilon_e(k_y)\phi_1(x)$$

We can take a guess at a solution that is localized near the boundary,

$$\phi_1(x) = e^{-\frac{1}{\lambda}\int_0^x dx' M(x')}\begin{pmatrix} a \\ b \end{pmatrix}$$

It can be verified as an eigenstate with eigenvalue $\varepsilon_e(k_y) = \lambda k_y$ if (a, b)=(1, i). Furthermore, it decays to zero at $x \ll 0$ and $x \gg 0$, and has a peak at x = 0. On the other hand, if

$$M(x)\begin{cases} > 0 \text{ for } x < 0 \\ < 0 \text{ for } x > 0 \end{cases}$$

then

$$\phi_1(x) = e^{-\frac{1}{\lambda}\int_0^x dx' M(x')}\begin{pmatrix} 1 \\ -i \end{pmatrix}$$

is a localized eigenstate with $\varepsilon_e(k_y) = -\lambda k_y$.

Therefore, in a sample with finite width, the electrons on the right edge move with velocity $\frac{1}{\hbar}\frac{\partial \varepsilon_e}{\partial k_y} = \lambda/\hbar$; the ones on the left move with velocity $-\lambda/\hbar$. They are called ***chiral edge states***. The two edges can be treated as independent only if the strip is wide enough (compared to the decay length of the edge state) so that the edge states on two sides do not couple with each other. In the small k_y limit, the energy dispersion $\varepsilon_e(k_y)$ of the edge states are linear in k_y.

The lattice Chern insulator – BHZ model

We now would like to make the simplest lattice generalization of the continuum Dirac Hamiltonian: instead of k_x, k_y, we generalize this to $d_1 = k_x \to \sin(k_x)$, $d_2 = k_y \to \sin(k_y)$, whereas the mass of the Dirac Hamiltonian can be generalized to $d_3 = m \to M(k) = -2B\left[2 - (M/2B) - \cos(k_x) - \cos(k_y)\right]$. The σ matrices now act in a space of orbitals. This lattice generalization becomes the continuum Dirac Hamiltonian for momentum close to $k = 0$. The lattice generalization of the Dirac Hamiltonian is an orbital model with two orbitals per site, one of *s*-type and one of *p*-type (or, more generally, orbitals of different parity). The fact that the coupling between them is odd in **k** means that they need to differ by one quantum of angular momentum, e.g., an *s*-type and a *p*-type orbital. Hence, the lowest-order coupling is linear in $\left(\sin(k_x) + i\sin(k_y)\right)$. We also add intraorbital dispersions $2 - \cos(k_x) - \cos(k_y)$ because they are allowed by symmetry.

Hence, to the lowest-order in Fourier modes, the Hamiltonian necessarily must look like

$$H = A\sin(k_x)\,\sigma_x + A\sin(k_y)\,\sigma_y - 2B\left[2 - (M/2B) - \cos(k_x) - \cos(k_y)\right]\sigma_z$$

The M term is a magnetization term that breaks time-reversal symmtry, and d_1, d_2 are used to decrib the spin-orbit coupling.

The eigen energies are

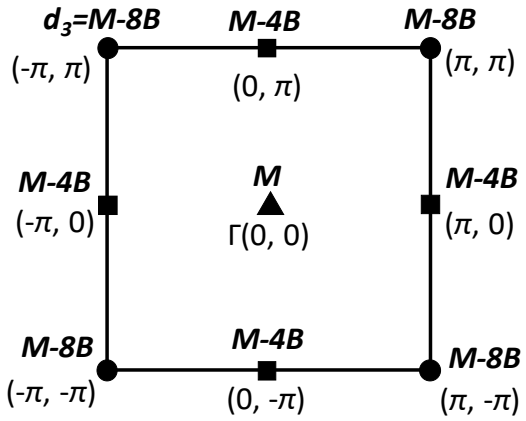
$$\varepsilon_{\pm}(\mathbf{k}) = \varepsilon_0(\mathbf{k}) \pm |\mathbf{h}(\mathbf{k})|$$

By $d_3 = 0$, it is not difficult to see that there are energy gaps at

$$\mathbf{k_0} = (0,0) \rightarrow \varepsilon_{\pm}(\mathbf{k_0}) = \varepsilon_0 \pm M$$

$$\mathbf{k_0} = (\pi,0),(0,\pi) \rightarrow \varepsilon_{\pm}(\mathbf{k_0}) = \varepsilon_0 \pm |M - 4B|$$

$$\mathbf{k_0} = (\pi,\pi) \rightarrow \varepsilon_{\pm}(\mathbf{k_0}) = \varepsilon_0 \pm |M - 8B|$$

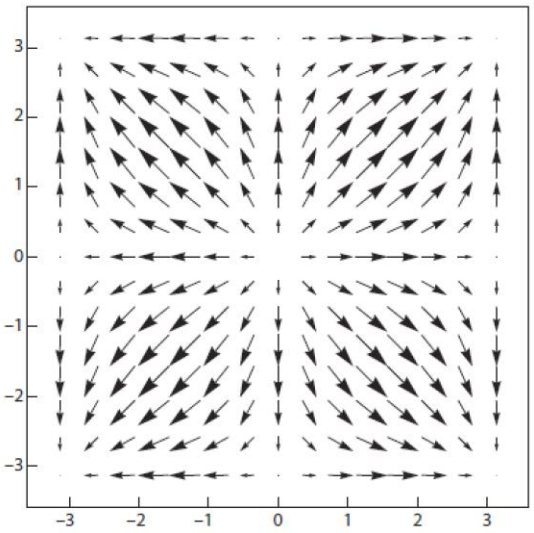


The energy gap closes at $M = 0, 4B, 8B$. We will see that the topology of the Bloch states changes at these critical points, when the energy gap closes. The distribution of the $h(\mathbf{k})$ vectors in the Brillouin zone changes when an energy gap closes.

By assuming $B < 0$, we have the following cases:

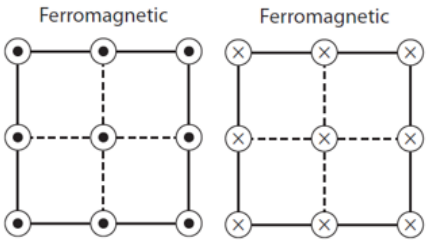
(i) $M > 0$ Phase and $M < 8B$ Phase

In these two regimes, the Hamiltonian is fully gapped and the Chern number is zero. Both $M > 0$ and $M < 8B$ represent Hamiltonians topologically equivalent to the atomic limit (the limit in which all hoppings are set to zero, as if the lattice constant were infinity). The $M > 0$ regime is topologically the same as the phase $M \rightarrow +\infty$, which has trivial eigenstates and zero Hall conductance. This is, in fact, an atomic limit, which does not show anything interesting because the energy bands are flat and only on-site energies are important so the wavefunctions are completely localized on the atomic sites. Here, $M < 8B$ is topologically the same as $M \rightarrow -\infty$, which is also an atomic limit but with the on-site energies reversed from the $M \rightarrow +\infty$ phase.



Vector plot for the components $d_1(k) = \sin(k_x)$, $d_2(k) = \sin(k_y)$ in the Chern insulator, $e_s = M/2B$.

Another way of understanding that the Chern number is zero is by seeing that the d_3 component of the Hamiltonian does not change sign in the BZ. The Hall conductance is a integrated over the whole BZs, which is identical to a skyrmion number of the vector $\hat{d}_i$. Because $\sin(k_x)$ and $\sin(k_y)$ form a vector that looks like the figure shown above, we see that to realize a skyrmion, we must look at the component of the d_3. The configuration of d_3 is ferromagnetic for the values of M in this region, and so there is no skyrmion in $\hat{d}_i$ and the spectrum is fully gapped, hence, no nonzero Hall conductance. i.e.



$$\sigma_{xy} = 0 \quad \text{for} \quad \frac{M}{2B} < 0 \quad \text{and} \quad \frac{M}{2B} > 4$$

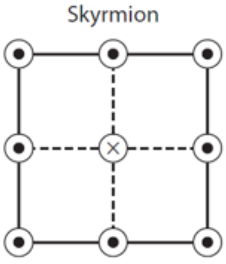
while in the former case, the "magnetic field" d_i only sweeps over the northern hemisphere when **k** cans over the BZ.

(ii) $4B < M < 0$ Phase

As we decrease M from the atomic limit ($M = +\infty$), at $M = 0$, the energy gap closes at the Γ = (0, 0) point. All other points remain gapped. Hence, we need to look only for the physics around that particular point if all we care about is the low-energy structure. We expand to obtain

$$H_{\Gamma+\mathbf{k}} = k_x\sigma_x + k_y\sigma_y + M\sigma_z$$

up to linear terms in k. At the gap-closing-and-reopening transition, M goes from positive to negative, so the change in Hall conductance is



$$\Delta\sigma_{xy} = \frac{e^2}{h}\left[\frac{1}{2}\text{sign}(M)_{M<0} - \frac{1}{2}\text{sign}(M)_{M>0}\right] = -1\frac{e^2}{h}$$

Because we know that the initial state had zero Hall conductance, by the preceding arguments, the new state has

$$\sigma_{xy} = -1\frac{e^2}{h}$$

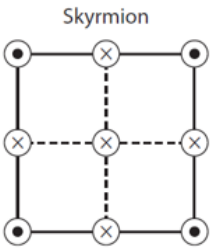
The Hall conductance corresponds to the skyrmion number in the BZ. For $4B < M < 0$, the configuration of d_3 looks like that in the figure above. As such, $d_3 < 0$ around the Γ point and then becomes positive at (π, 0), (0, π), where it becomes equal to $M - 4B$. This corresponds to a skyrmion number of −1. The difference between the two Hall conductances of the Dirac fermions is the skyrmion number equal to −1. i.e.

$$\sigma_{xy} = -1\frac{e^2}{h} \quad \text{for} \quad 0 < \frac{M}{2B} < 2$$

In this case, the "magnetic field" d_i changes sign and the gauge $\mathbf{A}_{North}$ is required to avoid the singularity. Therefore the topology is non-trivial and the Hall conductivity is non-zero.

(iii) $8B < M < 4B$ Phase

When we decrease M even further, we reach another phase transition. At $M = 4B$, the gap closes at two points in the BZ: (π, 0) and (0, π). We can analyze what happens at this phase transition by looking only around those two points. We linearize the Hamiltonian around these points:



$$H_{(\pi,0)+\mathbf{k}} = -k_x\sigma_x + k_y\sigma_y + (M-4B)\sigma_z$$

$$H_{(0,\pi)+\mathbf{k}} = k_x\sigma_x - k_y\sigma_y + (M-4B)\sigma_z$$

and we can see that the Hall conductivity of each of these Dirac Hamiltonians is

$$\sigma_{xy} = -\frac{1}{2}\text{sign}(M-4B)\frac{e^2}{h}$$

The overall '−' sign arises from the odd sign of the determinant A (See the next section). We hence have, for the transition between $M > 4B$ and $M < 4B$, a change in Hall conductance of

$$\Delta\sigma_{xy} = 2\left(-\frac{1}{2}\text{sign}(M-4B)_{M<4B} - \left(-\frac{1}{2}\text{sign}(M-4B)_{M>4B}\right)\right)\frac{e^2}{h} = 2\frac{e^2}{h}$$

Note that the factor '2' in front stems from the two gap closings. Because the phase before the transition ($M > 4B$) had Hall conductance −1, the new Hall conductance is

$$\sigma_{xy} = (-1+2)\frac{e^2}{h} = 1\frac{e^2}{h}$$

i.e. $\sigma_{xy} = 1\frac{e^2}{h}$ for $2 < \frac{M}{2B} < 4$

In this case, the "magnetic field" d_i changes sign again and the gauge $\mathbf{A}_{South}$ is required to avoid the singularity too. Therefore the topology is non-trivial and the Hall conductivity is non-zero.

(iv) Back to the Trivial State for $M < 8B$

At $M = 8B$, yet another point in the BZ becomes gapless: (π, π). Anywhere in between the values of $M = 0, 4B, 8B$, the model is gapped, as it is easily seen from the dispersion. Around the point (π, π), the Hamiltonian is

$$H_{(\pi,\pi)+\mathbf{k}} = -k_x\sigma_x - k_y\sigma_y + (M-8B)\sigma_z$$

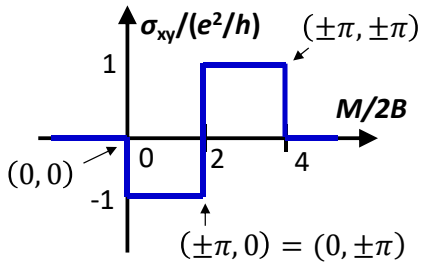
and the change in Hall conductance between $M > 8B$ and $M < 8B$ is

$$\Delta\sigma_{xy} = \left(\frac{1}{2}\text{sign}(M-8B)_{M<8B} - \frac{1}{2}\text{sign}(M-8B)_{M>8B}\right)\frac{e^2}{h} = -1\frac{e^2}{h}$$

which, when added to the previous Hall conductance, gives $\sigma_{xy} = 0$ for $M < 8B$. i.e.

$$\sigma_{xy} = 0 \quad \text{for} \quad \frac{M}{2B} > 4$$

This is similar to case (i), but the "magnetic field" d_i only sweeps over the southern hemisphere. The topology is again trivial.



We can now summarize our analysis. We started from a Hamiltonian that breaks time-reversal invariance $\mathcal{T}$ and found a non-vanishing σ_{xy}. We did so not by calculating $\mathcal{F}_{\mu v}$ and performing complicated **k**-space integrals but via a simple analysis of gap closings. The $\boldsymbol{h}$ is the "magnetic field" for the quasi-spin.

The simple picture presented above can be verified by actual calculation of σ_{xy} using the Berry phase:

The Berry curvature can be calculated as

$$\mathcal{F}_z^{\pm}(\mathbf{k}) = \mp\frac{1}{2h^3}\mathbf{h}\cdot\frac{\partial\mathbf{h}}{\partial k_x}\times\frac{\partial\mathbf{h}}{\partial k_y}$$

Proof: The Berry connections in **k**-space are

$$\mathcal{A}_l^{\pm}(\mathbf{k}) = i\langle\mathbf{h},\pm|\frac{\partial}{\partial k_l}|\mathbf{h},\pm\rangle = \frac{\partial h_\alpha}{\partial k_l}i\langle\mathbf{h},\pm|\frac{\partial}{\partial h_\alpha}|\mathbf{h},\pm\rangle = \frac{\partial h_\alpha}{\partial k_l}a_\alpha^{\pm}(\mathbf{h})$$

where $a_\alpha^{\pm}$ are the Berry connection in **h**–space. Therefore, the Berry curvatures in **k**-space are,

$$\mathcal{F}_z^{\pm}(\mathbf{k}) = \frac{\partial\mathcal{A}_y^{\pm}}{\partial k_x} - \frac{\partial\mathcal{A}_x^{\pm}}{\partial k_y} = \frac{\partial}{\partial k_x}\left(\frac{\partial h_\beta}{\partial k_y}a_\beta^{\pm}\right) - \frac{\partial}{\partial k_y}\left(\frac{\partial h_\alpha}{\partial k_x}a_\alpha^{\pm}\right)$$

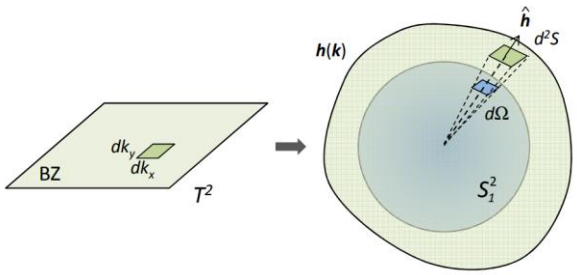
$$= \frac{\partial h_\alpha}{\partial k_x}\frac{\partial h_\beta}{\partial k_y}\left(\frac{\partial a_\beta^{\pm}}{\partial h_\alpha} - \frac{\partial a_\alpha^{\pm}}{\partial h_\beta}\right) = \frac{\partial h_\alpha}{\partial k_x}\frac{\partial h_\beta}{\partial k_y}\varepsilon_{\alpha\beta\gamma}b_\gamma^{\pm} = \mp\frac{1}{2h^3}\mathbf{h}\cdot\frac{\partial\mathbf{h}}{\partial k_x}\times\frac{\partial\mathbf{h}}{\partial k_y}$$

in which $b_\gamma^{\pm} = \mp\frac{h_\gamma}{2h^3}$ are the Berry curvatures in **h**-space. End of proof.

Suppose the lower band is completely filled, and the upper band is empty, then

$$\sigma_{xy} = \frac{e^2}{2\pi h}\int_{BZ} d^2k\, \mathcal{F}_z^-(\mathbf{k}) = \frac{e^2}{4\pi h}\int_{BZ} d^2k \frac{1}{h^3}\mathbf{h}\cdot\frac{\partial \mathbf{h}}{\partial k_x}\times\frac{\partial \mathbf{h}}{\partial k_y}$$

h describes a point on a two-dimensional sphere $\mathbf{S}^2$. This is the Bloch sphere. Specifically, in the integrand, $\hat{\mathbf{h}}\cdot\left(\frac{\partial \boldsymbol{h}}{\partial k_x}dk_x\right)\times\left(\frac{\partial \boldsymbol{h}}{\partial k_y}dk_y\right)$ is actually the area d^2S on the **h**-surface in Fig. 39. After being divided by h^2, it becomes the solid angle $d\Omega$ extended by that area. Since the BZ is a closed surface (a 2D torus), under a continuous mapping from $\mathbf{T}^2 \to \mathbf{S}^2$, it would map to a closed surface in **h**-space. In short, this formula measures the area of the unit sphere (counted with sign) swept out as we vary **k** over $\mathbf{T}^2$. In other words, it counts how many times $\mathbf{T}^2$ wraps around $\mathbf{S}^2$. The integral in σ_{xy} gives the total solid angle extended by that **h**-surface. For a closed surface, it must be an integer multiple of 4π, thus



Mapping a small area d^2k in 2D BZ to a small area d^2S on the surface of $\boldsymbol{h}(\mathbf{k})$. Its solid angle $d\Omega$ is equal to the area of d^2S projected on a unit sphere S_1^2.

$$\sigma_{xy} = \omega\frac{e^2}{h}, \omega \in \boldsymbol{Z}$$

The integer ω, which is equal to the first Chern number C_1, is the number of times the **h**-surface wraps over a unit sphere $\mathbf{S}^2$. It characterizes the topology of the mapping (and the Bloch states) and is called the winding number (or the wrapping number). We emphasize that ω is an integer only if the base space is a close surface ($\mathbf{T}^2$), which requires the valence band to be completely filled (an insulator).

The quantized Hall conductance in the lattice model is a result of the "magnetization" m, not an external magnetic field (as in the case of the quantum Hall effect). It is called alternatively as the quantum anomalous Hall effect (QAHE). Their difference is that, in the QHE, the electron orbitals are quantized due to the external magnetic field; in the QAHE there is no orbital quantization, but only spin re-orientation due to m.

Determinant Formula for the Hall Conductance of a Generic Dirac Hamiltonian

So far, we have obtained the Hall conductance for rotationally invariant Dirac Hamiltonians. We now obtain the Chern number for the generic Dirac Hamiltonian:

$$h(\mathbf{k}) = \sum_{a,b=1}^{2} k_a A_{ab}\sigma_b + M\sigma_3$$

where A_{ab} are numbers with $a, b = 1, 2$ (in a suitable r basis) and M is the Dirac gap. The field strength reads

$$\mathcal{F}_{12} = \frac{1}{2d^3} M(A_{22}A_{11} - A_{21}A_{12}) = \frac{1}{2d^3} M \det(A)$$

where d is described by $d^2 = k_m k_i A_{ij} A_{mj} + M^2$. Going to polar coordinates $k_x = k \sin\theta$, $k_y = k\cos\theta$, we write the Hall conductance as the integral over the occupied states of the Berry curvature; using the identity

$$\int_0^{2\pi} d\theta \frac{1}{\cos(\theta)^2 A_{1j}A_{1j} + \sin(\theta)^2 A_{2j}A_{2j} + 2\cos(\theta)\sin(\theta)A_{1j}A_{2j}} = \frac{2\pi}{|\det(A)|}$$

we have that the Hall conductance equals

$$\sigma_{xy} = \frac{1}{2}\frac{e^2}{h}\text{sign}(M)\text{sign}(\det(A))$$

Notice that the Hamiltonian $h(\mathbf{k})$ is not the most general Dirac Hamiltonian: this would be $k_\mu A_{\mu v}\sigma_v$, with $\mu, v = 1, 2, 3$ and $k_3 = M$. The Hall conductance of such a Hamiltonian is tedious but straightforward to obtain and equal $\text{sign}(\det(A))/2$.